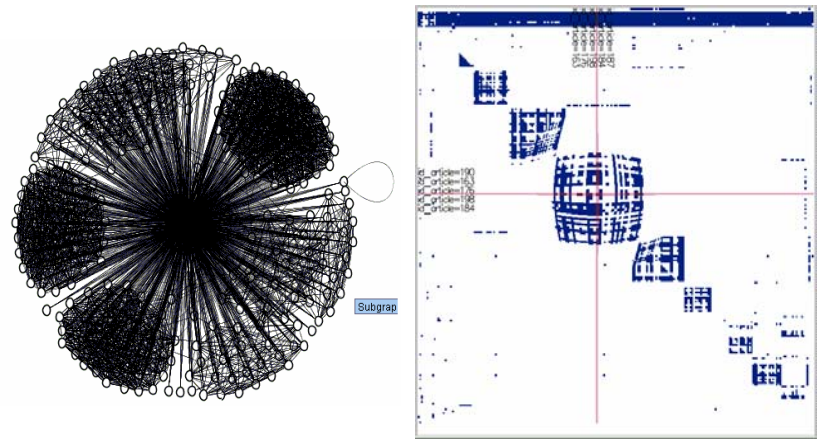
Visualizing Dense Networks with Enhanced and Hybrid Matrices

Jean-Daniel Fekete
AVIZ Team, INRIA Futurs
www.aviz.fr

INRIA FUTURS

Problem: Visualizing Dense Networks

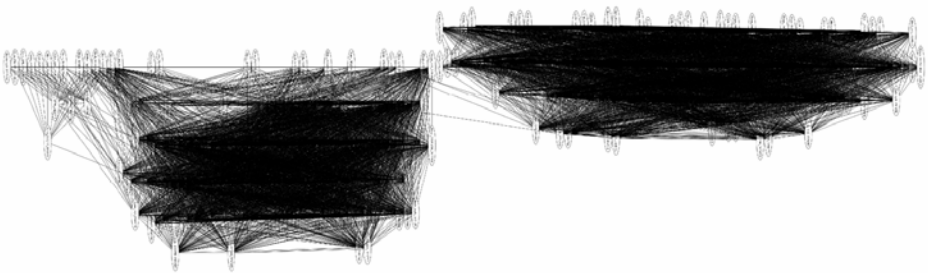


Example: Comparison of proteins of sequenced microbial genomes

- 570 microbial genomes sequenced (Jul. 07)
- 5 millions proteins
- Microbiogenomics project
  - Optimal extraction of relevant information from complex and heterogeneous data provided by exhaustive genomic comparisons
  - IGM+LRI/Univ. Paris-Sud, MIG/INRA
- Protein-Protein Network
  - Compute evolution distance
  - Cut above threshold (250 PAM units)
  - Cluster
  - Cut weak links (try to)
- Visualize!

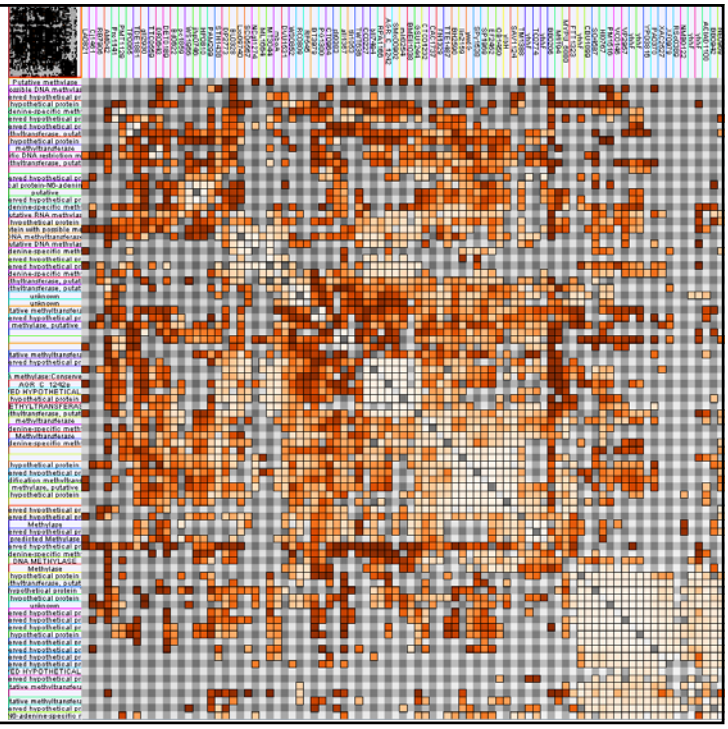
Node-Link Visualization of Protein Homology

- Two simple ANRt synthetase families
- Source: IGM/Univ. Paris-Sud (B. Labedan)



Name

Functional Annotation

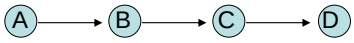


The Full Story

- What is an Adjacency Matrix?
- How does it compare with NL diagrams?
- How to order it?
- How to improve it?
- How does it scale?

Adjacency Matrix Construction

Directed Graph:



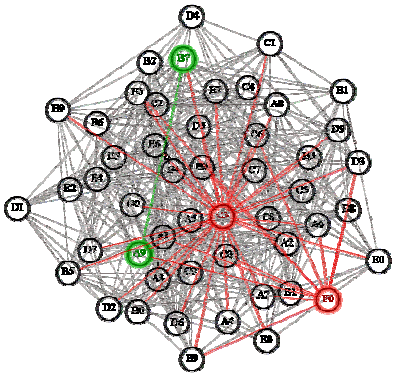
Adjacency Matrix:

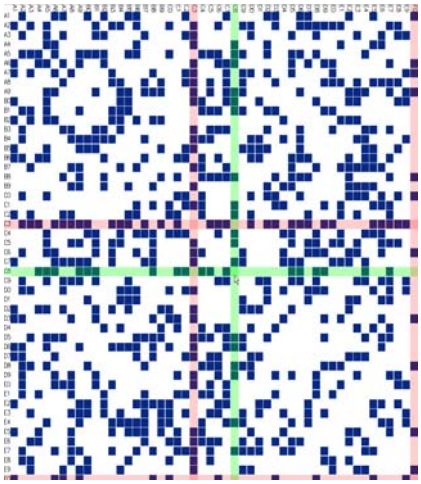
	A	B	C	D
A	0	1	0	0
B	0	0	1	0
C	0	0	0	1
D	0	0	0	0

To

From

Graphs Readability: Node Link Diagrams vs. Adjacency Matrices



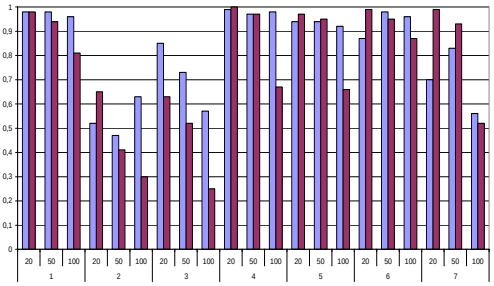


Comparing the readability of the 2 representations

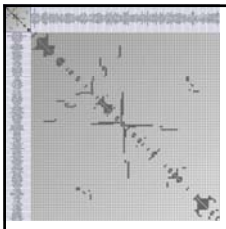
- **The Tasks:**
- Tasks related to the overview
  - Number of vertices
  - Number of arcs
- Tasks related to graph elements
  - Finding an element (a vertex, a link)
  - Finding the most connected vertex (a central actor, a pivot, a hub)
  - Finding a common neighbor
  - Finding a path
- Random graphs (3 sizes et 3 densities)
- 2 representations: Node-Link + Matrix
- **Results:**
- Node-link diagrams are preferable for small sparse graphs (20 vertices)
- Matrices are more readable *wrt* dense graphs and medium/large graphs (> 20 vertices) *wrt* the selected tasks, except paths

References:

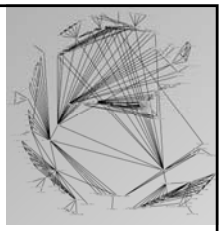
Mohammad Ghoniem, Jean-Daniel Fekete and Philippe Castagliola Readability of Graphs Using Node-Link and Matrix-Based Representations: Controlled Experiment and Statistical Analysis, Information Visualization Journal, 4(2), Palgrave Macmillan, Summer 2005, pp. 114-135.



Completion time for the 7 tasks, 3 densities and 2 representations (Node-Link in blue, Matrix in red)



Matrix vs. NodeLink



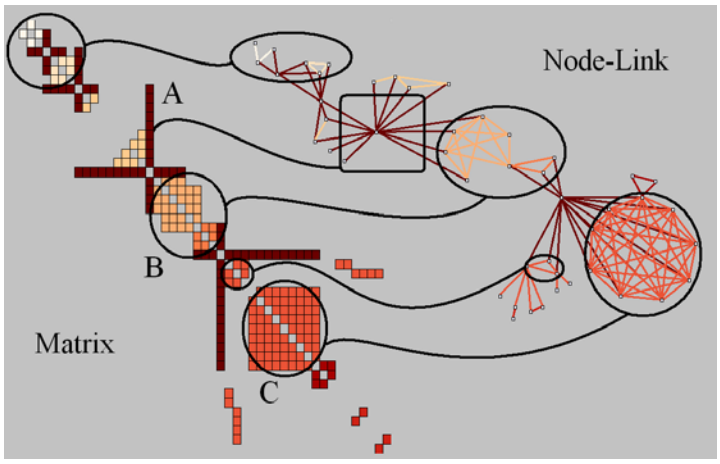
+

- Usable without reordering
- No node overlapping
  No edge crossing
  ➔ Readable for dense graphs
- Fast navigation
- Fast manipulation
  ➔ Usable interactively
- More readable for some tasks
- Familiar
- Compact
- More readable for path following
- More effective for small graphs
- More effective for sparse graphs

-

- Less familiar
- Use more space
- Weak for path following tasks
- Useless without layout
- Node overlapping
  Edge crossing
  ➔ Not readable for dense graphs
- Manipulation requires layout computation

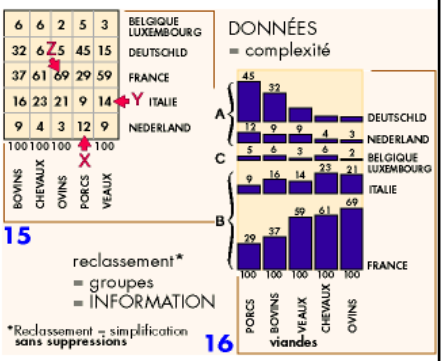
Visual Patterns with Ordered Matrices



The Reorderable Matrix

- Introduced by Bertin 67 as a representation for relational data
- Table or Network
- The value table provides details
- The reordered table provides details AND overall structure in the same representation

- Problems:
- how to compute a good ordering?
  - Row and column permutations
- how to assess its quality?

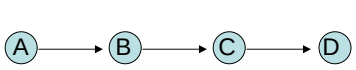


Reordering the Matrix

- Interactive or Automated
- Naïve approach:
  - Define an objective function (e.g. favor diagonal placement and dense clusters)
  - Try all permutations and retain the one that maximizes it
  - Problem : for a n×m table, there are n!×m! configurations
- Three families of methods to reorder tables:
  1. Robinsonian
  2. Dimension reduction
  3. Heuristics

Reordering Networks: Consider Adjacency Matrix as a Table

- Direct approach
  - Use the (weighted) adjacency table and reorder that table
- Enhanced approach (Henry & Fekete)
  - Use the (weighted) distance table and reorder that table



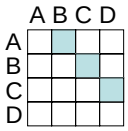
But

	A	B	C	D
A	0	1	0	0
B	0	0	1	0
C	0	0	0	1
D	0	0	0	0

Adjacency Matrix

$d_{A,B}$= 1, $d_{A,C}$= 1, $d_{A,D}$= 1

Should become



	A	B	C	D
A	x	1	2	3
B	1	x	1	2
C	2	1	x	1
D	3	2	1	x

Distance Matrix

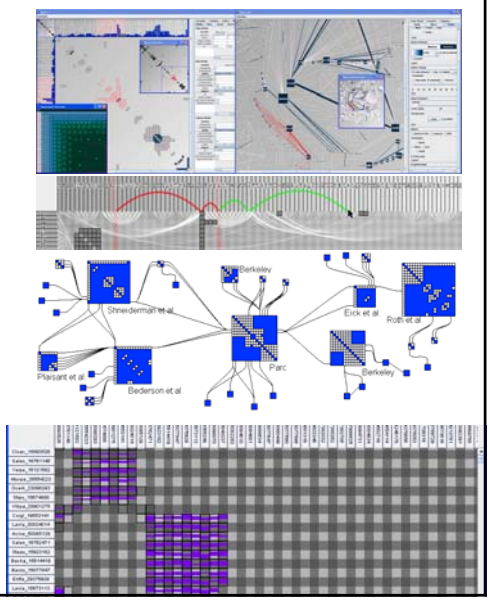
$d_{A,B}$= 4, $d_{A,C}$= 6, $d_{A,D}$= 8

What method is the best for me?

- Fast:
  - Transform into Table + PCA/CA (Harel&Koren 02)
  - Nearest-Neighbor TSP Heuristic
- Robust: Distance Table +
  - TSP
  - Clustering+seriation
  - Ellipse
- Others need more analysis

Improving Matrices

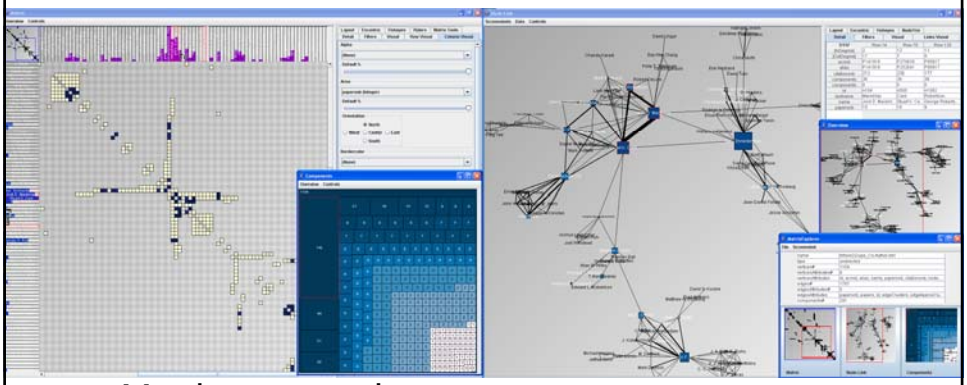
1. Dual Representation
   - MatrixExplorer (Henry&Fekete InfoVis06)
- Hybrid Visualization
  - MatLink (Henry&Fekete Interact07)
  - NodeTrix (Henry et al. InfoVis07)
- Multiscale Visualization
  - ZAME



1. Combine both representations

MatrixExplorer

MatrixExplorer [Henry&Fekete06]



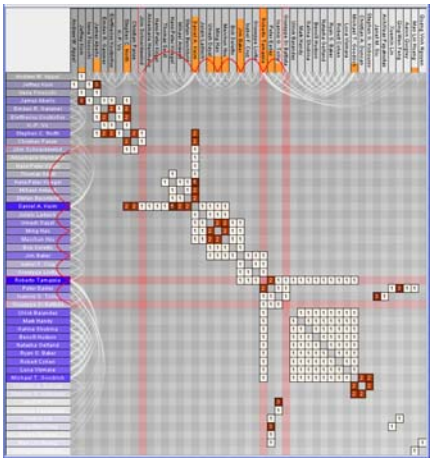
- Matrices to explore
- Node-Link diagrams to present findings

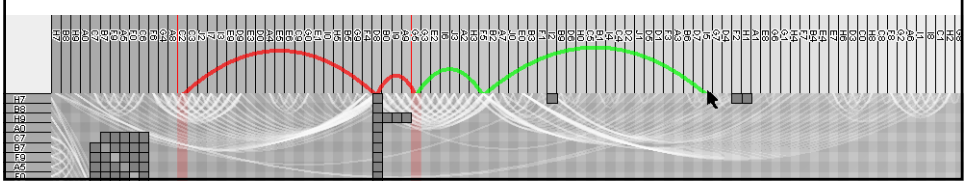
2. Augment one representation

MatLink

MatLink[Henry&Fekete07]

- Solving the path-related tasks problem for matrices
- Augmenting matrices with interactive links





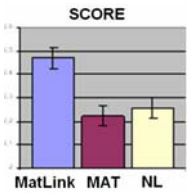
MatLink significantly improves matrices

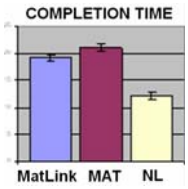
- Controlled experiment
  - 3 vis. x 6 datasets x 5 tasks

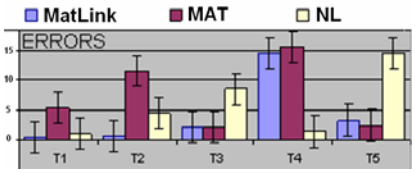
  Matrix , Node-Link, MatLink

  Data: From almost-trees
  To complete-graphs
  Including small-world networks

  Tasks:
  1. CommonNeighbour,
  2. ShortestPath,
  3. MostConnected,
  4. ArticulationPoint,
  5. LargestClique





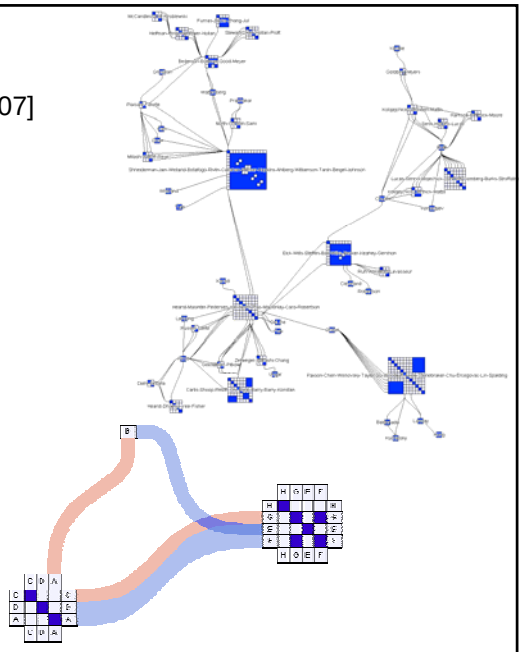


3. Find a hybrid representation

NodeTrix

NodeTrix[Henry et al.07]

- Designed for small-world networks
  - Globally sparse
  - Locally dense
- Visualizing dense sub-graphs as matrices
- Interact to create, edit and remove the matrices



NodeTrix

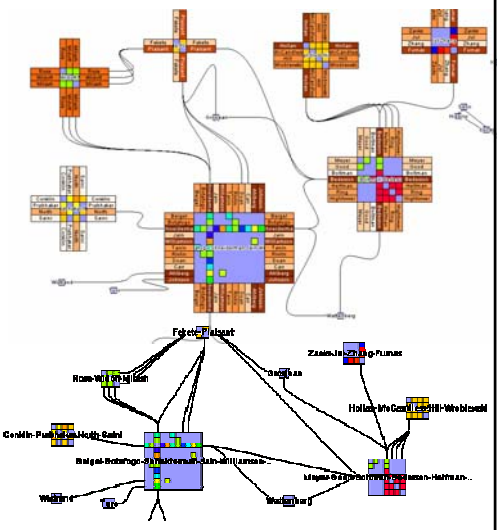
VIDEO at
http://insitu.lri.fr/~nhenry/nodetrix/nodetrix.mov

NodeTrix: the NetVis Nirvana?

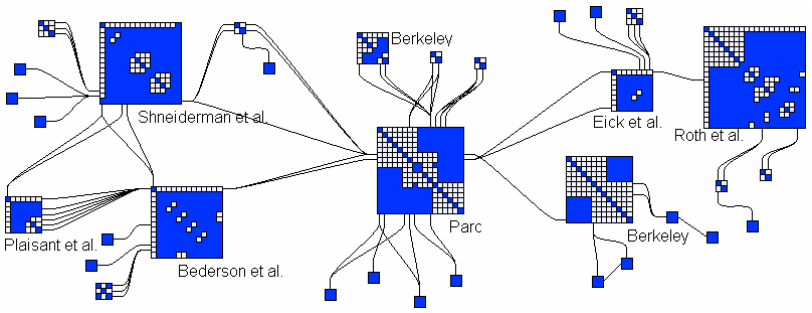
- ✓ Can you see every node?
- ✓ Can you count each node's degree?
- ✓ Can follow every link from its source to its destination?
- ✓ Can you identify clusters and outliers?

- Node Labels
- Link Labels (excentric labels?!)
- … even cluster labels
- Node Attributes
- Link Attributes
- … even clusters attributes
- Directed Graph (links width?!)

… But… beware the graphics overload!



Visual Patterns



Infovis Coauthorship (133 actors)

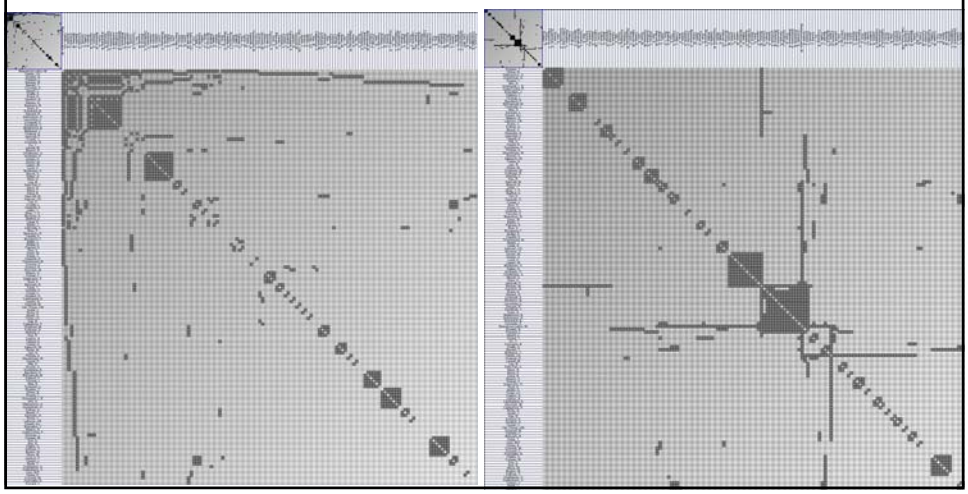
How does it scale?

- Visualize Wikipedia-Site Networks
  - 500 000 vertices
  - 6 000 000 edges
- Optimize reordering methods
- Use simple pyramid aggregation for topology
- Use sophisticated value aggregation for values
- Navigate by zooming using the wheel

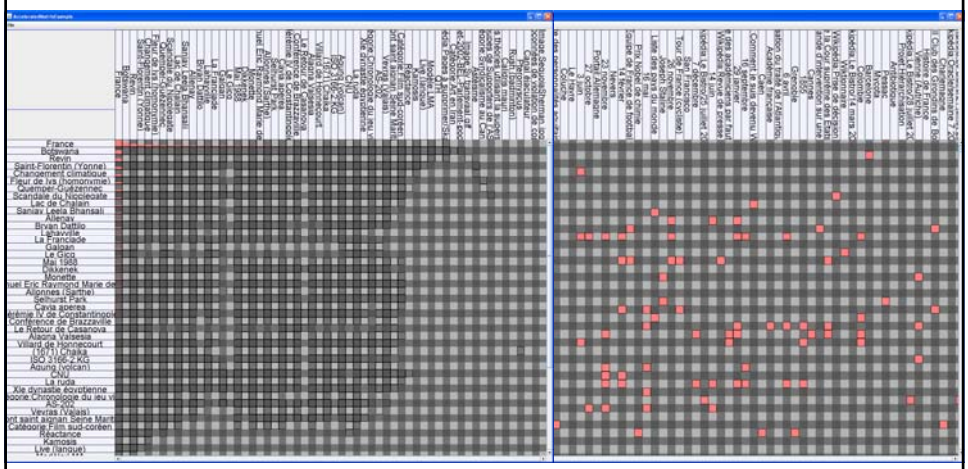
Non-Quadratic Methods

- PCA based dimension reduction methods
  - Harel&Koren “HDE” method
  - Generate good overviews and bad details
- Nearest-Neighbor Traveling Salesman Heuristics
  - Generate OK overviews and details
  - We tried several variations
    - Better are more expensive …

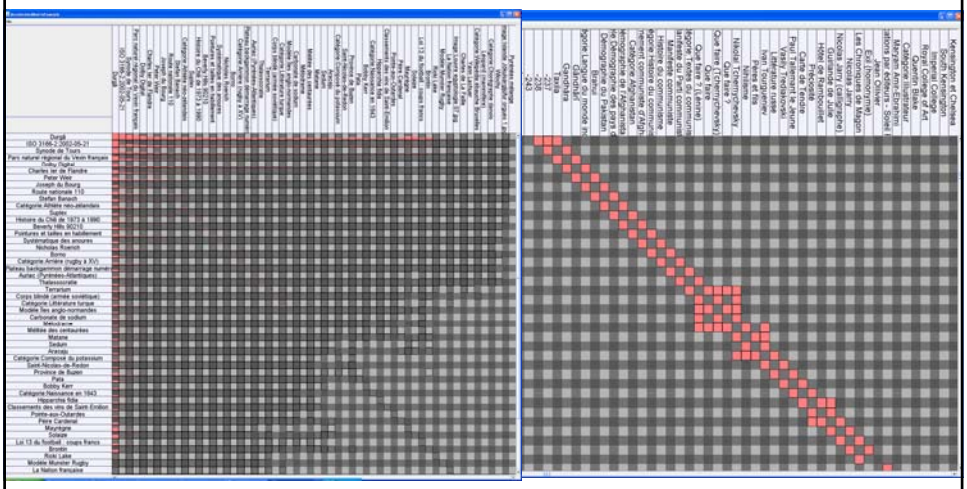
HDE vs. TSP based orderings



HDE overview and details

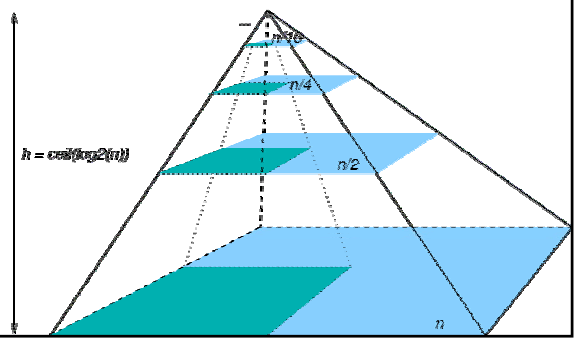


NNTSP overview and details



Topology Aggregation

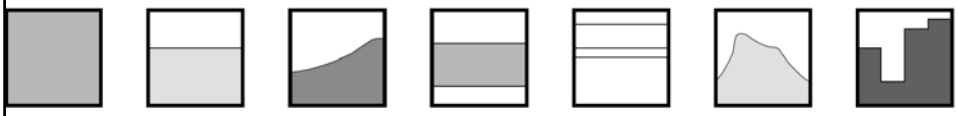
- Edges aggregated according to vertex order
- Data structure maintained to save memory
  - Row-column order
- Fast traversal
- Use paging to fit in memory
  - No limit!



Values Aggregation

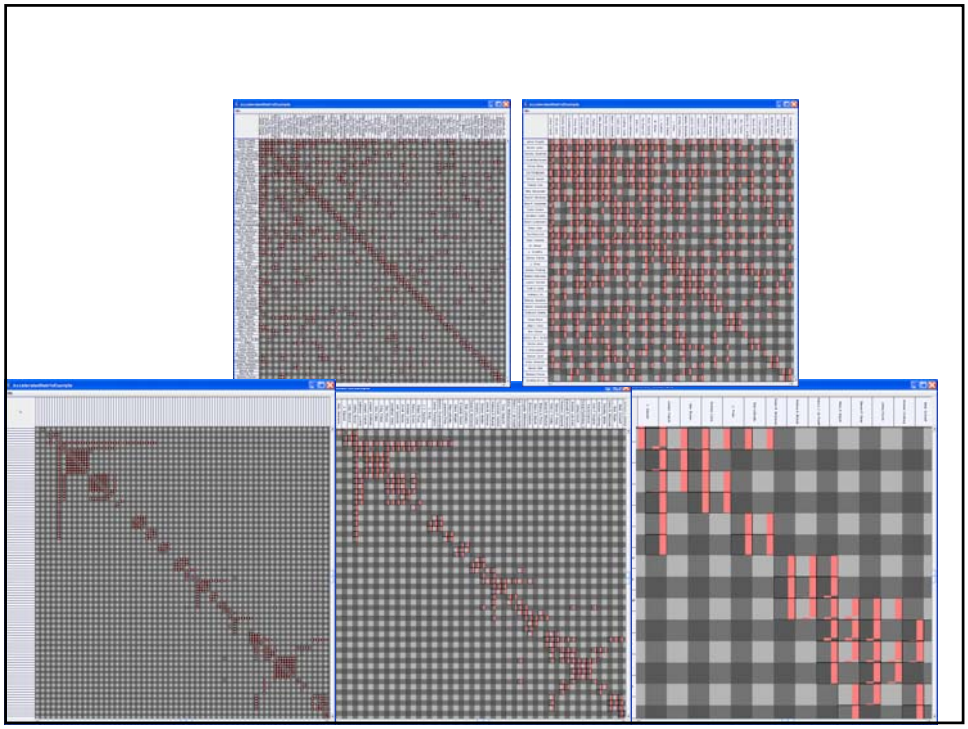
- Values can be:
  - Nominal, categorical (e.g. boolean), ordered or numerical
- Several ways to aggregate values
  - One -> one (Pivot tables or OLAP DB)
    - E.g. sample or concatenate nominal, count categorical, average numerical
  - One -> many : Symbolic Aggregation

Visual Representation of Aggregated Values



One-valued Interval-valued Histograms

- Problem:
  - Find a tradeoff between value precision, visual clutter and computation time
- Solution:
  - Count, min-max, 4 quantiles histogram



Results

Dataset	Vertices	Edges	Load (sec)	Reorder (sec)
InfoVis04	1,000	1,000	10	10
Protein-Protein	30,000	1,000,000	20	20
Wikipedia FR	500,000	6,000,000	65	70

Conclusion

- Matrix representation is a powerful complement to NL
- Good for
  - Dense networks
  - Filtering and selection
- Once filtered and reduced, the network can be visualized with a NL
- Hybrid representations
  - improve understanding of matrices
  - combine the best of both worlds

References

- N. Henry, J-D. Fekete and M. J. McGuffin, "NodeTrix: a Hybrid Visualization of Social Networks," IEEE Transactions on Visualization and Computer Graphics ,vol. 13, no. 6, pp. 1302-1309, November/December, 2007.
- N. Henry and J.-D. Fekete. *MatLink: Enhanced Matrix Visualization for Analyzing Social Networks.* In Cécilia Baranauskas, Philippe Palanque, Julio Abascal, and Simone Diniz Junqueira Barbosa, editors, Human-Computer Interaction – INTERACT 2007, volume 4663 of LNCS, pages 288–302. Springer, 2007. (Best paper)
- N. Henry and J.-D. Fekete. *MatrixExplorer: a Dual-Representation System to Explore Social Networks.* IEEE Transactions on Visualization and Computer Graphics (Proceedings Visualization / Information Visualization 2006), 12(5):677-684, September-October 2006.
- M. Ghoniem, J.-D. Fekete and P. Castagliola. *Readability of Graphs Using Node-Link and Matrix-Based Representations: Controlled Experiment and Statistical Analysis*. Information Visualization Journal, 4(2):114–135, 2005.

Acknowledgements

Thanh-Nghi Do
Niklas Elmqvist
Howard Goodell
Nathalie Henry